

MATS: Inference for potentially Singular and Heteroscedastic MANOVA

Sarah Friedrich* and Markus Pauly*

April 13, 2017

Abstract

In many experiments in the life sciences, several endpoints are recorded per subject. The analysis of such multivariate data is usually based on MANOVA models assuming multivariate normality and covariance homogeneity. These assumptions, however, are often not met in practice. Furthermore, test statistics should be invariant under scale transformations of the data, since the endpoints may be measured on different scales. In the context of high-dimensional data, Srivastava and Kubokawa (2013) proposed such a test statistic for a specific one-way model, which, however, relies on the assumption of a common non-singular covariance matrix. We modify and extend this test statistic to factorial MANOVA designs, incorporating general heteroscedastic models. In particular, our only distributional assumption is the existence of the group-wise covariance matrices, which may even be singular. We base inference on quantiles of resampling distributions, and derive confidence regions and ellipsoids based on these quantiles. In a simulation study, we extensively analyze the behavior of these procedures. Finally, the methods are applied to a cardiology study with unequal and singular covariance matrices.

Keywords: MANOVA, Multivariate Data, Parametric Bootstrap, Confidence Regions, Singular Covariance Matrices

1 Motivation and Introduction

In many experiments in the life sciences, several endpoints, potentially measured on different scales, are recorded per subject. As an example we consider a recent cardiology study conducted at the Department of Internal Medicine II, Ulm University, Germany, where 5 cardiological measurements were recorded in the left ventricle of 188 healthy patients. One aim is to analyze whether these variables differ between male and female patients. More precisely, the outcomes measured are peak systolic and diastolic strain rate (measured in $m\ell$) as well as end systolic and diastolic volume and stroke volume (in $1/sec$). In addition to unequal empirical covariance matrices between groups, analysis is further complicated by their singularity.

The analysis of such multivariate data is typically based on classical MANOVA models assuming multivariate normality and/or homogeneity of the covariance matrices, see, e.g., Lawley (1938); Bartlett (1939); Wilks (1946); Hotelling (1951); Pillai (1955); Dempster (1958, 1960). These assumptions, however, are often not met in practice (as in the motivating example) and it is well known that the methods perform poorly in case of heterogeneous data (Vallejo and Ato, 2012; Konietzschke et al., 2015). Furthermore, the test statistic should be invariant under scale transformations of the components, since the endpoints may be measured on different scales. Thus, test statistics of multivariate ANOVA-type (ATS) as, e.g., proposed in Brunner (2001) and studied in Friedrich et al. (2017), are only applicable if all endpoints are measured on the same scale, i.e., for repeated measures designs. Assuming non-singular covariance matrices and certain moment assumptions the scale invariance is typically accomplished by utilizing test statistics of Wald-type (WTS). However, inference procedures based thereon require (extremely) large sample sizes for being accurate, see Vallejo et al. (2010); Konietzschke et al. (2015); Smaga (2016). In particular, even the novel approaches of Konietzschke et al. (2015) and Smaga (2016) showed a more or less liberal behavior in case of skewed distributions. Moreover, their procedures cannot be used to analyze the motivating data example with possibly singular covariance matrices. Therefore, we follow a different approach by modifying the above mentioned ANOVA-type statistic (MATS). It is motivated from the modified Dempster statistic proposed in Srivastava and Kubokawa (2013) for high-dimensional one-way MANOVA. This statistic is also invariant under the change of units of measurements. However, until now, it has only been developed for a homoscedastic one-way setting assuming non-singularity and a specific distributional structure that is motivated from multivariate normality.

It is the aim of the present paper to modify and extend the Srivastava and Kubokawa (2013) approach to factorial MANOVA designs, incorporating general heteroscedastic models. In particular, we only postulate existence of the group-wise covariance matrices, which may even be singular. The small sample behavior of our test statistic is enhanced by applying bootstrap techniques as in Konietzschke et al. (2015). Thereby, the novel MATS procedure enables us to relax the usual MANOVA assumptions in several ways: While incorporating general heteroscedastic designs and allowing for potentially singular covariance matrices we postulate their existence as solely distributional assumption, i.e., only finite second moments are required. Moreover, we gain a procedure that is more robust against deviations from symmetry and homoscedasticity than the usual WTS approaches.

So far, only few approaches have been investigated which do not assume normality or equal covariance matrices (or both). Examples in the nonparametric framework include the permutation based nonparametric combination method (Pesarin and Salmaso, 2010, 2012) and the rank-based tests presented in Brunner et al. (1999) and Brunner et al. (2016) for Split Plot Designs and in Bathke et al. (2008) and Liu et al. (2011) for MANOVA designs. However, these methods are either not applicable for general MANOVA models or based on null hypotheses formulated in terms of distribution functions. In contrast we wish to derive inference procedures (tests and confidence regions) for contrasts of mean vectors. Here, beside all previously mentioned procedures, only methods for specific designs have been developed, see Chung and Romano (2016) for two-sample problems, Van Aelst and Willems (2011, 2013) for robust but homoscedastic one-way MANOVA or Harrar and Bathke (2012) for a particular two-way MANOVA. To our knowledge, mean-based MANOVA procedures allowing for possibly singular covariance matrices have not been developed so far.

The paper is organized as follows: In Section 2 we describe the statistical model and hypotheses. Furthermore, we propose a new test statistic, which is applicable to singular covariance matrices and is invariant under scale transformations of the data. In Section 3, we present two different resampling approaches, which are used for the derivation of statistical tests as well as confidence regions and simultaneous confidence intervals for contrasts in Section 4. The different approaches are compared in a large simulation study (Section 5), where we analyze different factorial designs with a wide variety of distributions and covariance settings. The motivating data example is analyzed in Section 6 and we conclude with some final remarks and discussion in Section 7. All proofs are deferred to the supplementary material, where we also provide further simulation results.

2 Statistical Model, Hypotheses and Test Statistics

Throughout the paper, we will use the following notation. We denote by \mathbf{I}_d the d -dimensional unit matrix and by \mathbf{J}_d the $d \times d$ matrix of 1's, i.e. $\mathbf{J}_d = \mathbf{1}_d \mathbf{1}_d'$, where $\mathbf{1}_d = (1, \dots, 1)'$ is the d -dimensional column vector of 1's. Furthermore, let $\mathbf{P}_d = \mathbf{I}_d - \frac{1}{d} \mathbf{J}_d$ denote the d -dimensional centering matrix. By \oplus and \otimes we denote the direct sum and the Kronecker product, respectively.

In order to cover different factorial designs of interest, we establish the general model

$$\mathbf{X}_{ik} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ik} \quad (2.1)$$

for treatment group $i = 1, \dots, a$ and individual $k = 1, \dots, n_i$, on which we measure d -variate observations. Here $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{id})' \in \mathbb{R}^d$ for $i = 1, \dots, a$. A factorial structure can be incorporated by splitting up indices, see e.g. Konietzschke et al. (2015). For fixed $1 \leq i \leq a$, the error terms $\boldsymbol{\epsilon}_{ik}$ are assumed to be independent and identically distributed d -dimensional random vectors, for which the following conditions hold:

$$(1) \ E(\boldsymbol{\epsilon}_{i1}) = \mathbf{0}, \ i = 1, \dots, a,$$

$$(2) \ 0 < \sigma_{is}^2 = \text{Var}(X_{iks}) < \infty, \ i = 1, \dots, a, \ s = 1, \dots, d,$$

$$(3) \ \text{Cov}(\epsilon_{i1}) = \mathbf{V}_i \geq 0, \ i = 1, \dots, a.$$

Thus, we only assume the existence of second moments. For convenience, we aggregate the individual vectors into $\mathbf{X} = (\mathbf{X}'_{11}, \dots, \mathbf{X}'_{an_a})'$ as well as $\boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_a)'$. Denote by $N = \sum_{i=1}^a n_i$ the total sample size. In order to derive asymptotic results in this framework, we will throughout assume the usual sample size condition:

$$n_i/N \rightarrow \kappa_i > 0, \ i = 1, \dots, a$$

as $N \rightarrow \infty$.

An estimator for $\boldsymbol{\mu}$ is given by the vector of pooled group means $\bar{\mathbf{X}}_{i\cdot} = \frac{1}{n_i} \sum_{k=1}^{n_i} \mathbf{X}_{ik}, i = 1, \dots, a$, which we denote by $\bar{\mathbf{X}}_{\bullet} = (\bar{\mathbf{X}}'_{1\cdot}, \dots, \bar{\mathbf{X}}'_{a\cdot})'$. The covariance matrix of $\sqrt{N} \bar{\mathbf{X}}_{\bullet}$ is given by

$$\boldsymbol{\Sigma}_N = \text{Cov}(\sqrt{N} \bar{\mathbf{X}}_{\bullet}) = \text{diag} \left(\frac{N}{n_i} \mathbf{V}_i : 1 \leq i \leq a \right),$$

where the group-specific covariance matrices \mathbf{V}_i are estimated by the empirical covariance matrices

$$\hat{\mathbf{V}}_i = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\mathbf{X}_{ik} - \bar{\mathbf{X}}_{i\cdot})(\mathbf{X}_{ik} - \bar{\mathbf{X}}_{i\cdot})'$$

resulting in

$$\hat{\boldsymbol{\Sigma}}_N = \text{diag} \left(\frac{N}{n_i} \hat{\mathbf{V}}_i : 1 \leq i \leq a \right).$$

In this semi-parametric framework, hypotheses are formulated in terms of the mean vector as $H_0 : \mathbf{H}\boldsymbol{\mu} = \mathbf{0}$, where \mathbf{H} is a suitable contrast matrix, i.e., $\mathbf{H}\mathbf{1}_{ad} = \mathbf{0}$. Note that we can use the unique projection matrix $\mathbf{T} = \mathbf{H}'(\mathbf{H}\mathbf{H}')^+ \mathbf{H}$, where $(\mathbf{H}\mathbf{H}')^+$ denotes the Moore-Penrose inverse of $\mathbf{H}\mathbf{H}'$. It is $\mathbf{T} = \mathbf{T}^2, \mathbf{T} = \mathbf{T}'$ and $\mathbf{T}\boldsymbol{\mu} = \mathbf{0} \Leftrightarrow \mathbf{H}\boldsymbol{\mu} = \mathbf{0}$, see e.g. Brunner and Puri (2001).

A commonly used test statistic for multivariate data is the Wald-type statistic (WTS)

$$T_N = N \bar{\mathbf{X}}_{\bullet}' \mathbf{T} (\mathbf{T} \hat{\boldsymbol{\Sigma}}_N \mathbf{T})^+ \mathbf{T} \bar{\mathbf{X}}_{\bullet}, \quad (2.2)$$

which requires the additional assumption (3') $\mathbf{V}_i > 0, i = 1, \dots, a$. It is easy to show that the WTS has under $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, asymptotically, as $N \rightarrow \infty$, a χ_f^2 -distribution with $f = \text{rank}(\mathbf{T})$ degrees of freedom if (1) – (3') holds. However, large sample sizes are necessary to maintain a pre-assigned level α using quantiles of the limiting χ^2 -distribution. Konietschke et al. (2015) proposed different resampling procedures in order to improve the small sample behavior of the

WTS for multivariate data. Therein, a parametric bootstrap approach turned out to be the best in case that the underlying distributions are not too skewed and/or too heteroscedastic. In the latter cases all considered procedures were more or less liberal. Moreover, assuming only (3) instead of (3') for the WTS would in general not lead to an asymptotic χ_f^2 -limit distribution. In this paper we not only want to relax the assumption (3') on the unknown covariance matrices but also like to gain a procedure that is more robust to deviations from symmetry and homoscedasticity. To this end, we consider a different test statistic, namely a multivariate version of the ANOVA-type statistic (ATS) proposed by Brunner (2001) for repeated measures designs, which we obtain by erasing the Moore-Penrose term from (2.2):

$$\tilde{Q}_N = N\bar{\mathbf{X}}_{\bullet}'\mathbf{T}\bar{\mathbf{X}}_{\bullet}.$$

In the special two-sample case where we wish to test the null hypothesis $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$, this is equivalent to the test statistic proposed by Dempster (1960).

The drawback of the ATS for multivariate data is that it is not invariant under scale transformations of the components, e.g., under change of units ($cm \mapsto m$ or $g \mapsto kg$) in one or more components. We therefore consider a slightly modified version of the ATS, which we denote as MATS:

$$Q_N = N\bar{\mathbf{X}}_{\bullet}'\mathbf{T}(\mathbf{T}\hat{\mathbf{D}}_N\mathbf{T})^+\mathbf{T}\bar{\mathbf{X}}_{\bullet}. \quad (2.3)$$

Here, $\hat{\mathbf{D}}_N = \text{diag}\left(\frac{N}{n_i}\hat{\sigma}_{is}^2\right)$, $i = 1, \dots, a$, $s = 1, \dots, d$, where $\hat{\sigma}_{11}^2, \dots, \hat{\sigma}_{ad}^2$ are the diagonal entries of $\hat{\boldsymbol{\Sigma}}_N$, i.e., $\hat{\sigma}_{is}^2$ is the empirical variance of component s in group i . A related test statistic has been proposed by Srivastava and Kubokawa (2013) in the context of high-dimensional ($d \rightarrow \infty$) data for a special non-singular one-way MANOVA design. Here, we investigate in the classical multivariate case (with fixed d) how its finite sample performance can be enhanced considerably. We nevertheless start by analyzing its asymptotic limit behavior.

THEOREM 2.1 *Under $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, the test statistic Q_N in (2.3) has asymptotically, as $N \rightarrow \infty$, the same distribution as a weighted sum of χ_1^2 distributed random variables, where the weights λ_{is} are the eigenvalues of $\mathbf{V} = \mathbf{T}(\mathbf{T}\mathbf{D}\mathbf{T})^+\mathbf{T}\boldsymbol{\Sigma}$ for $\mathbf{D} = \text{diag}(\kappa_i^{-1}\sigma_{is}^2)$ and $\boldsymbol{\Sigma} = \text{diag}(\kappa_i^{-1}\mathbf{V}_i)$, i.e.,*

$$Q_N = N\bar{\mathbf{X}}_{\bullet}'\mathbf{T}(\mathbf{T}\hat{\mathbf{D}}_N\mathbf{T})^+\mathbf{T}\bar{\mathbf{X}}_{\bullet} \xrightarrow{\mathcal{D}} Z = \sum_{i=1}^a \sum_{s=1}^d \lambda_{is} Z_{is},$$

with $Z_{is} \stackrel{i.i.d.}{\sim} \chi_1^2$ and " $\xrightarrow{\mathcal{D}}$ " denoting convergence in distribution. Thus, we obtain a test $\varphi_N = \mathbb{1}\{Q_N > c_{1-\alpha}\}$, where $c_{1-\alpha}$ is the $(1 - \alpha)$ -quantile of Z .

In order to approximate the unknown limiting distribution and to derive adequate and asymptotically correct inference procedures based on Q_N in (2.3), we consider two different bootstrap approaches, which will be explained in detail in the next section.

Apart from statistical test decisions discussed in Section 4.1, a central part of statistical analyses is the construction of confidence intervals, which allows for deeper insight into the variability and the magnitude of effects. This univariate concept can be generalized to multivariate endpoints by constructing multivariate confidence regions and simultaneous confidence intervals for contrasts $\mathbf{c}'\boldsymbol{\mu}$ for any contrast vector $\mathbf{c} \in \mathbb{R}^{ad}$ of interest. Details on the derivation of such confidence regions for $\mathbf{c}'\boldsymbol{\mu}$ are given in Section 4.2 below.

3 Bootstrap Procedures

The first bootstrap procedure we consider is a parametric bootstrap approach as proposed by Konietzschke et al. (2015) for the WTS. The second one is a wild bootstrap approach, which has already been successfully applied in the context of repeated measures or clustered data, see Cameron et al. (2008); Cameron and Miller (2015) or Friedrich et al. (2016). Both bootstrap approaches are based on the test statistic Q_N in (2.3). Note that the procedures derived in the following can also be used for multiple testing problems, either by applying the closed testing principle (Marcus et al., 1976; Sonnemann, 2008) or in the context of simultaneous contrast tests (Hothorn et al., 2008; Hasler and Hothorn, 2008).

3.1 A Parametric Bootstrap Approach

This asymptotic model based bootstrap approach has successfully been used in univariate one- and two-way factorial designs (Krishnamoorthy and Lu (2010) or Xu et al. (2013)), and has recently been applied to Wald-type statistics for general MANOVA by Konietzschke et al. (2015) and Smaga (2016), additionally assuming finite fourth moments. The approach is as follows: Given the data, we generate a parametric bootstrap sample as

$$\mathbf{X}_{i1}^*, \dots, \mathbf{X}_{in_i}^* \stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \widehat{\mathbf{V}}_i), \quad i = 1, \dots, a.$$

The idea behind this method is to obtain an accurate finite sample approximation by mimicking the covariance structure given in the observed data. This is achieved by calculating Q_N^* from the bootstrap variables $\mathbf{X}_{i1}^*, \dots, \mathbf{X}_{in_i}^*$, i.e.,

$$Q_N^* = N(\overline{\mathbf{X}}_\bullet^*)' \mathbf{T}(\mathbf{T} \widehat{\mathbf{D}}_N^* \mathbf{T})^+ \mathbf{T} \overline{\mathbf{X}}_\bullet^*. \quad (3.1)$$

We then obtain a parametric bootstrap test by comparing the original test statistic Q_N with the conditional $(1 - \alpha)$ -quantile $c_{1-\alpha}^*$ of its bootstrap version Q_N^* .

THEOREM 3.1 *The conditional distribution of Q_N^* converges weakly to the null distribution of Q_N in probability for any parameters $\boldsymbol{\mu} \in \mathbb{R}^{ad}$ and $\boldsymbol{\mu}_0$ with $\mathbf{T}\boldsymbol{\mu}_0 = \mathbf{0}$. In particular,*

$$\sup_{x \in \mathbb{R}} |P_{\boldsymbol{\mu}}(Q_N^* \leq x | \mathbf{X}) - P_{\boldsymbol{\mu}_0}(Q_N \leq x)| \xrightarrow{P} 0,$$

where $P_{\boldsymbol{\mu}}(Q_N \leq x)$ and $P_{\boldsymbol{\mu}}(Q_N^* \leq x | \mathbf{X})$ denote the unconditional and conditional distribution of Q_N and Q_N^* , respectively, if $\boldsymbol{\mu}$ is the true underlying mean vector.

3.2 A Wild Bootstrap Approach

Another resampling approach, which is based on multiplying the fixed data with random weights, is the so-called wild bootstrap procedure. To this end, let W_{ik} denote i.i.d. random variables, independent of \mathbf{X} , with $E(W_{ik}) = 0$, $\text{Var}(W_{ik}) = 1$ and $\sup_{i,k} E(W_{ik}^4) < \infty$. We obtain a bootstrap sample as

$$\mathbf{X}_{ik}^* = W_{ik}(\mathbf{X}_{ik} - \overline{\mathbf{X}}_{i\cdot}), i = 1, \dots, a, k = 1, \dots, n_i.$$

Note that there are different choices for the random weights W_{ik} , e.g., Rademacher distributed random variables (Davidson and Flachaire, 2008) or weights satisfying different moment conditions, see, e.g., Wu (1986); Liu (1988); Mammen (1993); Beyersmann et al. (2013). The choice of the weights typically depends on the situation. In our simulation studies, we have investigated the performance of different weights such as Rademacher distributed as well as $N(0, 1)$ distributed weights (see Lin (1997) for this specific choice). The results were rather similar and we therefore only present the results of our simulation study for standard normally distributed weights in Section 5 below.

Based on the bootstrap variables \mathbf{X}_{ik}^* , we can calculate Q_N^* in the same way as described for Q_N^* in (3.1) above. A wild bootstrap test is finally obtained by comparing Q_N to the conditional $(1 - \alpha)$ -quantile of its wild bootstrap version Q_N^* .

THEOREM 3.2 *The conditional distribution of Q_N^* converges weakly to the null distribution of Q_N in probability for any parameters $\boldsymbol{\mu} \in \mathbb{R}^{ad}$ and $\boldsymbol{\mu}_0$ with $\mathbf{T}\boldsymbol{\mu}_0 = \mathbf{0}$. In particular,*

$$\sup_{x \in \mathbb{R}} |P_{\boldsymbol{\mu}}(Q_N^* \leq x | \mathbf{X}) - P_{\boldsymbol{\mu}_0}(Q_N \leq x)| \xrightarrow{P} 0.$$

4 Statistical Applications

We now want to base statistical inference on the modified test statistic in (2.3) using the two bootstrap approaches described above. A thorough statistical analysis should ideally consist of two parts: First, statistical tests give insight into significant effects of the different factors as well as possible interactions. We therefore consider important properties of statistical tests based on the bootstrap approaches in Section 4.1. Second, it is helpful to construct confidence regions for the unknown parameters of interest in order to gain a more detailed insight into the nature of the estimates. The derivation of such confidence regions is discussed in Section 4.2.

4.1 Statistical Tests

In this section, we analyze the statistical properties of the two bootstrap procedures described above. For ease of notation, we will only state the results for the parametric bootstrap procedure,

i.e., consider the test statistic Q_N^* based on \mathbf{X}_{ik}^* throughout. Note, however, that the results are also valid for the wild bootstrap procedure, i.e., the test statistic Q_N^* .

As mentioned above, a bootstrap test $\varphi^* = \mathbb{1}\{Q_N > c_{1-\alpha}^*\}$ is obtained by comparing the original test statistic Q_N to the $(1 - \alpha)$ -quantile $c_{1-\alpha}^*$ of its bootstrap version. In particular, p -values are numerically computed as follows:

- (1) Given the data \mathbf{X} , calculate the MATS Q_N for the null hypothesis of interest.
- (2) Bootstrap the data with either of the two bootstrap approaches described above and calculate the corresponding test statistic $Q_N^{*,1}$.
- (3) Repeat step (2) a large number of times, e.g. $B = 10,000$ times, and obtain values $Q_N^{*,1}, \dots, Q_N^{*,B}$.
- (4) Calculate the p -value based on the empirical distribution of $Q_N^{*,1}, \dots, Q_N^{*,B}$ as

$$p\text{-value} = \frac{1}{B} \sum_{b=1}^B \mathbb{1}\{Q_N \leq Q_N^{*,b}\}.$$

Theorems 3.1 and 3.2 imply that the corresponding tests asymptotically keep the pre-assigned level α under the null hypothesis and are consistent for any fixed alternative $\mathbf{T}\boldsymbol{\mu} \neq \mathbf{0}$, i.e., $E_{\boldsymbol{\mu}}(\varphi^*) \rightarrow \alpha \cdot \mathbb{1}\{\mathbf{T}\boldsymbol{\mu} = \mathbf{0}\} + \mathbb{1}\{\mathbf{T}\boldsymbol{\mu} \neq \mathbf{0}\}$. Moreover, for local alternatives $H_1 : \mathbf{T}\boldsymbol{\mu} = \frac{1}{\sqrt{N}}\mathbf{T}\boldsymbol{\nu}, \boldsymbol{\nu} \in \mathbb{R}^{ad}$, the bootstrap tests have the same asymptotic power as $\varphi_N = \mathbb{1}\{Q_N > c_{1-\alpha}\}$, where $c_{1-\alpha}$ is the $(1 - \alpha)$ -quantile of Z given in Theorem 2.1. In particular, the asymptotic relative efficiency of the bootstrap tests compared to φ_N is 1 in this situation.

4.2 Confidence regions and confidence intervals for contrasts

In order to conduct a thorough statistical analysis, interpretation of the results should not be based on p -values alone. In addition, it is helpful to construct confidence regions for the unknown parameter. The concept of a confidence region is the same as that of a confidence interval in the univariate setting: We want to construct a multivariate region, which is likely to contain the true, but unknown parameter of interest. The aim of this section is to derive multivariate confidence regions and simultaneous confidence intervals for contrasts $\mathbf{c}'\boldsymbol{\mu}$ for any contrast vector \mathbf{c} of interest. Such contrasts include, e.g., the difference in means $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ in two-sample problems, Dunnett's many-to-one comparisons, Tukey's all-pairwise comparisons, and many more, see, e.g., Hothorn et al. (2008) for specific examples. In this section, we will base the derivation of confidence regions on the bootstrap approximations given in Section 3, i.e., we will use one of the bootstrap quantiles. Again, we only formulate the results for $c_{1-\alpha}^*$.

Consider contrasts $\mathbf{C}\boldsymbol{\mu}, \mathbf{C} = (\mathbf{c}'_1, \dots, \mathbf{c}'_q) \in \mathbb{R}^{q \times ad}$. By the result from Section 4.1 it follows, that the null hypothesis for a single contrast $H_0^\ell : \mathbf{c}'_\ell \boldsymbol{\mu} = 0, \ell \in \{1, \dots, q\}$ is rejected, if the test statistic

$$Q_N^\ell = N(\mathbf{c}'_\ell \bar{\mathbf{X}}_\bullet)'(\mathbf{c}'_\ell \hat{\mathbf{D}}_N \mathbf{c}_\ell)^+(\mathbf{c}'_\ell \bar{\mathbf{X}}_\bullet)$$

is larger than the corresponding bootstrap quantile $c_{1-\alpha}^*$. Thus, a simultaneous $100(1 - \alpha)\%$ confidence interval for a single contrast $\mathbf{c}'_\ell \boldsymbol{\mu}$ is given by

$$\mathbf{c}'_\ell \bar{\mathbf{X}}_\bullet \pm \sqrt{c_{1-\alpha}^* \cdot \mathbf{c}'_\ell \hat{\mathbf{D}}_N \mathbf{c}_\ell / N}.$$

Consequently, a confidence region for contrasts $\mathbf{C}\boldsymbol{\mu}$ is determined by the set of all $\mathbf{C}\boldsymbol{\mu}$ such that

$$Q_N = N(\mathbf{C}\bar{\mathbf{X}}_\bullet - \mathbf{C}\boldsymbol{\mu})'(\mathbf{C}\hat{\mathbf{D}}_N\mathbf{C}')^+(\mathbf{C}\bar{\mathbf{X}}_\bullet - \mathbf{C}\boldsymbol{\mu}) \leq c_{1-\alpha}^*,$$

see the detailed derivations in Konietzschke et al. (2012) for univariate rank-based procedures as well as in Johnson and Wichern (2007, Chapter 6) for multivariate normally distributed data, which directly carry over to the present situation. Alternatively, we may apply the bootstrap procedure to obtain asymptotically valid simultaneous contrast tests formulated in the usual maximum statistic.

Furthermore, a confidence ellipsoid is obtained based on the eigenvalues $\hat{\lambda}_s$ and eigenvectors $\hat{\mathbf{e}}_s$ of $\mathbf{C}\hat{\mathbf{D}}_N\mathbf{C}'$. As in Johnson and Wichern (2007), the direction and lengths of the axes of

$$N(\mathbf{C}\bar{\mathbf{X}}_\bullet - \mathbf{C}\boldsymbol{\mu})'(\mathbf{C}\hat{\mathbf{D}}_N\mathbf{C}')^+(\mathbf{C}\bar{\mathbf{X}}_\bullet - \mathbf{C}\boldsymbol{\mu}) \leq c_{1-\alpha}^*$$

are determined by going $\sqrt{\hat{\lambda}_s \cdot c_{1-\alpha}^* / N}$ units along the eigenvectors $\hat{\mathbf{e}}_s$ of $\mathbf{C}\hat{\mathbf{D}}_N\mathbf{C}'$. In other words, the axes of the ellipsoid are

$$\mathbf{C}\bar{\mathbf{X}}_\bullet \pm \sqrt{\hat{\lambda}_s \cdot c_{1-\alpha}^* / N} \cdot \hat{\mathbf{e}}_s, \quad s = 1, \dots, d. \quad (4.1)$$

While we can calculate (4.1) for arbitrary dimensions d , we cannot display the joint confidence region graphically for $d \geq 4$.

In the two-sample case with $d = 2$ endpoints, however, the ellipse can be plotted: Beginning at the center $\mathbf{C}\bar{\mathbf{X}}_\bullet$, the axes of the ellipsoid are given by $\pm \sqrt{\hat{\lambda}_s \cdot c_{1-\alpha}^* / N} \cdot \hat{\mathbf{e}}_s$, $s = 1, 2$. That is, the confidence ellipse extends $\sqrt{\hat{\lambda}_s \cdot c_{1-\alpha}^* / N}$ units along the estimated eigenvector $\hat{\mathbf{e}}_s$ for $s = 1, 2$. Therefore, we get a graphical representation of the relation between the group-mean differences $\mu_{11} - \mu_{21}$ and $\mu_{12} - \mu_{22}$ of the first and second component, see Section 6 and Figure 3 below for an example.

5 Simulations

The procedures described in Section 3 are valid for large sample sizes. In order to investigate their behavior for small samples, we have conducted various simulations. In the simulation studies, the behavior of the proposed approaches was compared to a parametric bootstrap approach for the WTS as in Konietzschke et al. (2015) since this turned out to perform better than other resampling versions of the WTS and Wilk's Λ . All simulations were conducted using R Version 3.3.1 (R Core Team, 2016) each with $\text{nsim} = 5,000$ simulation and $\text{nboot} = 5,000$ bootstrap runs. We investigated a one- and a two-factorial design. Furthermore, we simulated data similar to the cardiology example, see the supplementary material for details and results.

5.1 One-way layout

For the one-way layout, data was generated as in Konietzschke et al. (2015). We considered $a = 2$ treatment groups and $d \in \{4, 8\}$ endpoints as well as the following covariance settings:

$$\begin{aligned} \text{Setting 1:} \quad & \mathbf{V}_1 = \mathbf{I}_d + 0.5(\mathbf{J}_d - \mathbf{I}_d) = \mathbf{V}_2, \\ \text{Setting 2:} \quad & \mathbf{V}_1 = ((0.6)^{|r-s|})_{r,s=1}^d = \mathbf{V}_2, \\ \text{Setting 3:} \quad & \mathbf{V}_1 = \mathbf{I}_d + 0.5(\mathbf{J}_d - \mathbf{I}_d) \text{ and } \mathbf{V}_2 = \mathbf{I}_d \cdot 3 + 0.5(\mathbf{J}_d - \mathbf{I}_d), \\ \text{Setting 4:} \quad & \mathbf{V}_1 = ((0.6)^{|r-s|})_{r,s=1}^d \text{ and } \mathbf{V}_2 = ((0.6)^{|r-s|})_{r,s=1}^d + \mathbf{I}_d \cdot 2. \end{aligned}$$

Setting 1 represents a compound symmetry structure, while setting 2 is an autoregressive covariance structure. Both settings 1 and 2 represent homoscedastic scenarios while settings 3 and 4 display two scenarios with unequal covariance structures. Data was generated by

$$\mathbf{X}_{ik} = \boldsymbol{\mu}_i + \mathbf{V}_i^{1/2} \boldsymbol{\epsilon}_{ik}, \quad i = 1, \dots, a; \quad k = 1, \dots, n_i,$$

where $\mathbf{V}_i^{1/2}$ denotes the square root of the matrix \mathbf{V}_i , i.e., $\mathbf{V}_i = \mathbf{V}_i^{1/2} \cdot \mathbf{V}_i^{1/2}$. The i.i.d. random errors $\boldsymbol{\epsilon}_{ik} = (\epsilon_{ik1}, \dots, \epsilon_{ikd})'$ with mean $E(\boldsymbol{\epsilon}_{ik}) = \mathbf{0}_d$ and $\text{Cov}(\boldsymbol{\epsilon}_{ik}) = \mathbf{I}_{d \times d}$ were generated by simulating independent standardized components

$$\epsilon_{iks} = \frac{Y_{iks} - E(Y_{iks})}{\sqrt{\text{Var}(Y_{iks})}}$$

for various distributions of Y_{iks} . In particular, we simulated normal, χ_3^2 , lognormal, t_3 and double-exponential distributed random variables. We investigated balanced as well as unbalanced designs with sample size vectors $\mathbf{n}^{(1)} = (10, 10)$, $\mathbf{n}^{(2)} = (20, 20)$, $\mathbf{n}^{(3)} = (10, 20)$ and $\mathbf{n}^{(4)} = (20, 10)$, respectively. A major criterion concerning the accuracy of the procedures is their behavior in situations where increasing variances (settings 3 and 4 above) are combined with increasing sample sizes ($\mathbf{n}^{(3)}$, positive pairing) or decreasing sample sizes ($\mathbf{n}^{(4)}$, negative pairing).

In this setting, we tested the null hypothesis $H_0^\mu : \{(\mathbf{P}_a \otimes \mathbf{I}_d)\boldsymbol{\mu} = \mathbf{0}\} = \{\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2\}$, i.e., no treatment effect. The resulting type-I error rates (nominal level $\alpha = 5\%$) for $d = 4$ and $d = 8$ endpoints are displayed in Table 1 (normal distribution) and Table 2 (χ_3^2 -distribution), respectively. Further simulation results for lognormal, t_3 and double-exponential distributed errors can be found in Tables ?? – ?? in the supplementary material.

As we can see, the MATS approach with the parametric bootstrap approximation works very well for most scenarios, keeping the pre-assigned level of 5% very accurately. Especially for $d = 8$ endpoints, it is often closer to the nominal level than the parametric bootstrap of the WTS. The wild bootstrap of the MATS, in contrast, shows a rather conservative behavior, especially with $d = 8$ endpoints and normally distributed data.

Especially in situations with negative pairing, see e.g. covariance settings 3 and 4 with $\mathbf{n}^{(4)} = (20, 10)$ in Table 1, the MATS approach with the parametric bootstrap works much better than the WTS parametric bootstrap. It also seems that these tendencies are more pronounced for large d , i.e., in situations where d is close to $\min(n_1, n_2)$.

However, in situations with negative pairing and skewed distributions (see Table 2 as well as Table ?? in the supplementary material), the parametric bootstrap MATS shows a slightly liberal behavior. For t_3 and double-exponentially distributed errors and negative pairing, in contrast, the parametric bootstrap MATS is slightly conservative, see Tables ?? and ?? in the supplementary material, respectively.

Surprisingly, both MATS approaches improve with growing d in most settings, i.e., when the number of endpoints is closer to the sample size. The WTS approach, in contrast, gets worse in these scenarios. This might be an interesting approach for future research in high-dimensional settings such as in Pauly et al. (2015).

As a result, we find that the MATS with the parametric bootstrap approximation is the best procedure in most scenarios. Especially, it is less conservative than the wild bootstrap approximation and less liberal than the WTS equipped with the parametric bootstrap approach over all simulation settings. Only in situations with negative pairing and skewed distributions, the new procedure shows a slightly liberal behavior. However, in these cases all considered procedures show considerable deviations from the nominal level α .

Table 1: Type-I error rates in % (nominal level $\alpha = 5\%$) for the parametric bootstrap (PBS) of the WTS and the MATS, as well as the wild bootstrap of the MATS in the one-way layout for the normal distribution.

d	Cov	n	WTS (PBS)	MATS (wild)	MATS (PBS)
$d = 4$	S1	(10, 10)	4.2	5.4	4.6
		(10, 20)	5.5	6.5	5.4
		(20, 10)	5.5	6.5	5.4
		(20, 20)	5.1	7.1	4.9
	S2	(10, 10)	4.1	5.3	4.6
		(10, 20)	5.5	6.3	5.0
		(20, 10)	5.6	6.2	5.0
		(20, 20)	5.1	7.1	5.1
	S3	(10, 10)	5.2	3.4	4.2
		(10, 20)	5.0	5.0	4.6
		(20, 10)	6.9	4.3	4.7
		(20, 20)	5.0	4.9	4.5
	S4	(10, 10)	5.2	3.5	4.0
		(10, 20)	4.9	5.0	4.4
		(20, 10)	7.0	4.2	4.8
		(20, 20)	5.0	4.8	4.6
$d = 8$	S1	(10, 10)	4.2	2.4	4.9
		(10, 20)	6.1	4.6	5.2
		(20, 10)	6.3	4.6	5.3
		(20, 20)	4.5	5.8	4.3
	S2	(10, 10)	4.3	1.3	4.5
		(10, 20)	6.1	3.0	4.3
		(20, 10)	6.6	3.5	5.1
		(20, 20)	4.7	4.6	4.1
	S3	(10, 10)	6.0	0.5	3.5
		(10, 20)	4.8	3.0	4.5
		(20, 10)	11.3	0.6	3.9
		(20, 20)	5.6	2.8	3.8
	S4	(10, 10)	5.9	0.4	3.6
		(10, 20)	4.7	2.0	4.1
		(20, 10)	11.5	0.5	3.7
		(20, 20)	5.7	2.2	4.0

Table 2: Type-I error rates in % (nominal level $\alpha = 5\%$) for the parametric bootstrap (PBS) of the WTS and the MATS, as well as the wild bootstrap of the MATS in the one-way layout for the χ_3^2 -distribution.

d	Cov	n	WTS (PBS)	MATS (wild)	MATS (PBS)
$d = 4$	S1	(10, 10)	3.9	4.4	3.9
		(10, 20)	5.1	7.4	5.8
		(20, 10)	5.2	7.1	5.2
		(20, 20)	4.0	7.3	4.9
	S2	(10, 10)	4.0	4.5	4.0
		(10, 20)	5.3	7.3	5.8
		(20, 10)	5.2	6.8	5.1
		(20, 20)	3.9	7.2	4.7
	S3	(10, 10)	6.6	4.8	5.0
		(10, 20)	4.3	5.2	4.5
		(20, 10)	10.3	8.0	8.3
		(20, 20)	6.4	6.2	5.5
	S4	(10, 10)	6.5	4.8	5.1
		(10, 20)	4.4	5.1	4.5
		(20, 10)	10.2	8.2	8.4
		(20, 20)	6.4	6.1	5.5
$d = 8$	S1	(10, 10)	3.6	2.4	4.5
		(10, 20)	5.8	5.4	5.6
		(20, 10)	5.3	5.1	5.1
		(20, 20)	4.3	6.8	4.7
	S2	(10, 10)	3.5	1.3	3.9
		(10, 20)	5.8	3.3	4.8
		(20, 10)	5.3	3.0	4.3
		(20, 20)	4.3	5.0	4.5
	S3	(10, 10)	6.2	1.4	5.1
		(10, 20)	4.9	3.5	4.7
		(20, 10)	13.1	3.2	7.7
		(20, 20)	6.1	4.2	5.1
	S4	(10, 10)	6.3	1.0	5.1
		(10, 20)	5.0	2.4	4.1
		(20, 10)	13.0	2.8	7.6
		(20, 20)	6.1	3.7	5.5

5.1.1 Singular Covariance Matrix

In order to analyze the behavior of the discussed methods in designs involving singular covariance matrices, we considered the one-way layout described above with $a = 2$ groups and

$d \in \{4, 8\}$ observations involving the following covariance settings (displayed for $d = 4$):

$$\begin{aligned}
\text{Setting 5: } \mathbf{V}_1 &= \begin{pmatrix} 1 & 1/2 & 1 & 1 \\ 1/2 & 1 & 1/2 & 1/2 \\ 1 & 1/2 & 1 & 1 \\ 1 & 1/2 & 1 & 1 \end{pmatrix}, \mathbf{V}_2 = \mathbf{V}_1 + 0.5 \cdot \mathbf{J}_d \\
\text{Setting 6: } \mathbf{V}_1 &= \begin{pmatrix} 1 & 0.6 & 0.36 & 0.18 \\ 0.6 & 1 & 0.6 & 0.3 \\ 0.36 & 0.6 & 1 & 0.5 \\ 0.18 & 0.3 & 0.5 & 0.25 \end{pmatrix}, \mathbf{V}_2 = \mathbf{V}_1 + 0.5 \cdot \mathbf{J}_d \\
\text{Setting 7: } \mathbf{V}_1 &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0.5 \end{pmatrix}, \mathbf{V}_2 = \mathbf{V}_1 + 0.5 \cdot \mathbf{J}_d
\end{aligned}$$

Setting 6 is based on an $AR(0.6)$ covariance matrix (see setting 2 above), where the last row and column have been replaced by half the row/column before, respectively. Setting 7 is based on $\tilde{\mathbf{V}}_1 = \text{diag}(\sqrt{2^s})$, $s = 0, \dots, d-1$, where the last row and column have been replaced by half the row/column before. We have considered sample sizes $\mathbf{n}^{(1)}$, $\mathbf{n}^{(3)}$ and $\mathbf{n}^{(4)}$ in order to cover a balanced design as well as designs with negative and positive pairing.

The results are displayed in Tables 3 and 4. The parametric bootstrap of the MATS again yields the best results. It is very accurate for all simulated scenarios. In contrast, the wild bootstrap MATS surprisingly shows a liberal behavior in covariance setting 5 and leads to very conservative decisions in setting 7 for $d = 8$. The parametric bootstrap WTS is rather liberal in case of the χ_3^2 -distribution and in settings 6 and 7 with $\mathbf{n}^{(4)}$ (negative pairing).

Finally, note that the WTS approach does not yield a valid level α test in this situation due to the singular covariance matrices.

5.2 Two-way layout

To analyze the behavior of our methods in a setting with two crossed factors A and B , we considered the following simulation design, which is again adapted from Konietzschke et al. (2015). We simulated a 2×2 design with sample sizes $\mathbf{n}^{(1)} = (n_{11}^{(1)}, n_{12}^{(1)}, n_{21}^{(1)}, n_{22}^{(1)}) = (10, 10, 10, 10)$, $\mathbf{n}^{(2)} = (7, 10, 13, 16)$, $\mathbf{n}^{(3)} = (16, 13, 10, 7)$, $\mathbf{n}^{(4)} = (20, 20, 20, 20)$. The covariance settings were chosen similar to the one-way layout above as:

Table 3: Type-I error rates in % (nominal level $\alpha = 5\%$) for the parametric bootstrap (PBS) of the WTS and the MATS, as well as the wild bootstrap of the MATS in the one-way layout with singular covariance matrices for the normal distribution.

d	Cov	n	WTS (PBS)	MATS (wild)	MATS (PBS)
$d = 4$	S5	(10, 10)	4.8	12.2	4.7
		(10, 20)	5.1	12.7	5.0
		(20, 10)	5.0	12.9	5.2
	S6	(10, 10)	4.5	6.8	4.7
		(10, 20)	4.8	7.4	4.9
		(20, 10)	6.1	8.2	5.4
	S7	(10, 10)	4.6	4.6	3.9
		(10, 20)	4.7	5.1	4.4
		(20, 10)	6.1	5.8	5.2
$d = 8$	S5	(10, 10)	4.5	17.4	5.0
		(10, 20)	5.1	18.7	5.0
		(20, 10)	5.3	18.8	5.2
	S6	(10, 10)	3.8	2.5	4.9
		(10, 20)	5.8	3.6	4.4
		(20, 10)	6.9	5.2	5.8
	S7	(10, 10)	3.7	0.6	3.2
		(10, 20)	5.6	1.5	3.8
		(20, 10)	7.1	2.1	4.1

Setting 8: $\mathbf{V}_{ij} = \mathbf{I}_d + 0.5(\mathbf{J}_d - \mathbf{I}_d), i, j = 1, 2,$

Setting 9: $\mathbf{V}_{ij} = ((0.6)^{|r-s|})_{r,s=1}^d, i, j = 1, 2,$

Setting 10: $\mathbf{V}_{ij} = \mathbf{I}_d \cdot \ell + 0.5(\mathbf{J}_d - \mathbf{I}_d), \ell = 1, \dots, 4,$

Setting 11: $\mathbf{V}_{ij} = ((0.6)^{|r-s|})_{r,s=1}^d + \mathbf{I}_d \cdot \ell, \ell = 1, \dots, 4.$

Again, setting 10 and 11 combined with sample sizes $\mathbf{n}^{(2)}$ and $\mathbf{n}^{(3)}$ represent settings with positive and negative pairing, respectively. In this scenario, we consider three different null hypotheses of interest:

- (1) The hypothesis of *no effect of factor A*

$$H_0^\mu(A) : \{\bar{\boldsymbol{\mu}}_{\cdot 1} = \bar{\boldsymbol{\mu}}_{\cdot 2}\} = \{\mathbf{H}_A \boldsymbol{\mu} = \mathbf{0}\},$$

- (2) The hypothesis of *no effect of factor B*

$$H_0^\mu(B) : \{\bar{\boldsymbol{\mu}}_{\cdot 1} = \bar{\boldsymbol{\mu}}_{\cdot 2}\} = \{\mathbf{H}_B \boldsymbol{\mu} = \mathbf{0}\},$$

Table 4: Type-I error rates in % (nominal level $\alpha = 5\%$) for the parametric bootstrap (PBS) of the WTS and the MATS, as well as the wild bootstrap of the MATS in the one-way layout with singular covariance matrices for the χ_3^2 -distribution.

d	Cov	n	WTS (PBS)	MATS (wild)	MATS (PBS)
$d = 4$	S5	(10, 10)	4.0	11.7	4.1
		(10, 20)	4.9	13.3	5.3
		(20, 10)	5.6	13.5	5.5
	S6	(10, 10)	7.5	6.3	4.4
		(10, 20)	7.1	7.8	5.4
		(20, 10)	9.3	8.8	5.8
	S7	(10, 10)	7.3	4.8	4.2
		(10, 20)	7.0	5.4	4.5
		(20, 10)	9.2	6.6	5.8
$d = 8$	S5	(10, 10)	4.3	17.1	4.5
		(10, 20)	5.1	18.3	4.8
		(20, 10)	5.3	17.9	5.1
	S6	(10, 10)	5.8	2.6	4.7
		(10, 20)	7.0	4.4	4.7
		(20, 10)	8.4	5.6	5.5
	S7	(10, 10)	5.7	0.4	3.2
		(10, 20)	6.8	1.6	3.7
		(20, 10)	8.3	2.2	4.6

(3) The hypothesis of *no* $A \times B$ interaction effect

$$H_0^\mu(AB) : \{(\mathbf{P}_a \otimes \mathbf{P}_b \otimes \mathbf{I}_d)\boldsymbol{\mu} = \mathbf{0}\},$$

where $\mathbf{H}_A = \mathbf{P}_a \otimes \frac{1}{b}\mathbf{J}_b \otimes \mathbf{I}_d$ and $\mathbf{H}_B = \frac{1}{a}\mathbf{J}_a \otimes \mathbf{P}_b \otimes \mathbf{I}_d$.

The results for the interaction hypothesis are displayed in Table 5. Since the total sample size N is larger in this scenario, the asymptotic results come into play and therefore all methods lead to more accurate results than in the one-way layout. We find a similar behavior as in the one-way layout: Again, the MATS and the WTS with the parametric bootstrap approach control the type-I error very accurately, whereas the wild bootstrap approach leads to slightly liberal test decisions in this setting. The difference between symmetric and skewed distributions is not as pronounced here as it was in the one-way layout. When testing for an effect of factor A or B in situations with negative pairing (covariance setting 10 and 11 with sample size vector $\mathbf{n}^{(3)}$), the parametric bootstrap MATS improves the slightly liberal behavior of the WTS, see Tables ?? and ?? in the supplementary material, respectively. Note that we have only considered $d = 4$ endpoints in the two-way layout.

Table 5: Type-I error rates in % (nominal level $\alpha = 5\%$) for the parametric bootstrap (PBS) of the WTS and the MATS, as well as the wild bootstrap of the MATS when testing the interaction hypothesis in a two-way layout with $d = 4$ dimensional observations.

distr	Cov	n	WTS (PBS)	MATS (wild)	MATS (PBS)
normal	S8	$\mathbf{n}^{(1)}$	4.8	6.9	4.9
		$\mathbf{n}^{(2)}$	5.1	6.7	4.6
		$\mathbf{n}^{(3)}$	5.1	7.3	5.2
		$\mathbf{n}^{(4)}$	4.8	7.7	5.1
	S9	$\mathbf{n}^{(1)}$	4.8	6.5	4.6
		$\mathbf{n}^{(2)}$	5.0	6.7	4.8
		$\mathbf{n}^{(3)}$	5.0	7.1	5.3
		$\mathbf{n}^{(4)}$	4.7	7.7	5.2
	S10	$\mathbf{n}^{(1)}$	4.7	4.8	4.6
		$\mathbf{n}^{(2)}$	4.6	5.2	4.7
		$\mathbf{n}^{(3)}$	5.6	5.0	4.9
		$\mathbf{n}^{(4)}$	4.7	5.3	4.8
	S11	$\mathbf{n}^{(1)}$	4.9	4.4	4.3
		$\mathbf{n}^{(2)}$	4.6	4.8	4.7
		$\mathbf{n}^{(3)}$	5.3	4.8	4.8
		$\mathbf{n}^{(4)}$	4.9	4.8	4.6
χ_3^2	S8	$\mathbf{n}^{(1)}$	3.9	7.1	4.9
		$\mathbf{n}^{(2)}$	4.4	6.9	4.6
		$\mathbf{n}^{(3)}$	4.6	7.0	4.4
		$\mathbf{n}^{(4)}$	4.4	7.3	4.7
	S9	$\mathbf{n}^{(1)}$	3.8	6.8	4.4
		$\mathbf{n}^{(2)}$	4.3	6.6	4.6
		$\mathbf{n}^{(3)}$	4.6	6.9	4.6
		$\mathbf{n}^{(4)}$	4.4	7.4	4.7
	S10	$\mathbf{n}^{(1)}$	4.0	4.2	3.5
		$\mathbf{n}^{(2)}$	4.0	4.9	4.0
		$\mathbf{n}^{(3)}$	4.7	4.4	3.9
		$\mathbf{n}^{(4)}$	4.7	4.8	4.1
	S11	$\mathbf{n}^{(1)}$	3.9	4.0	3.5
		$\mathbf{n}^{(2)}$	4.1	4.4	3.9
		$\mathbf{n}^{(3)}$	4.7	4.0	3.7
		$\mathbf{n}^{(4)}$	4.6	4.6	4.2

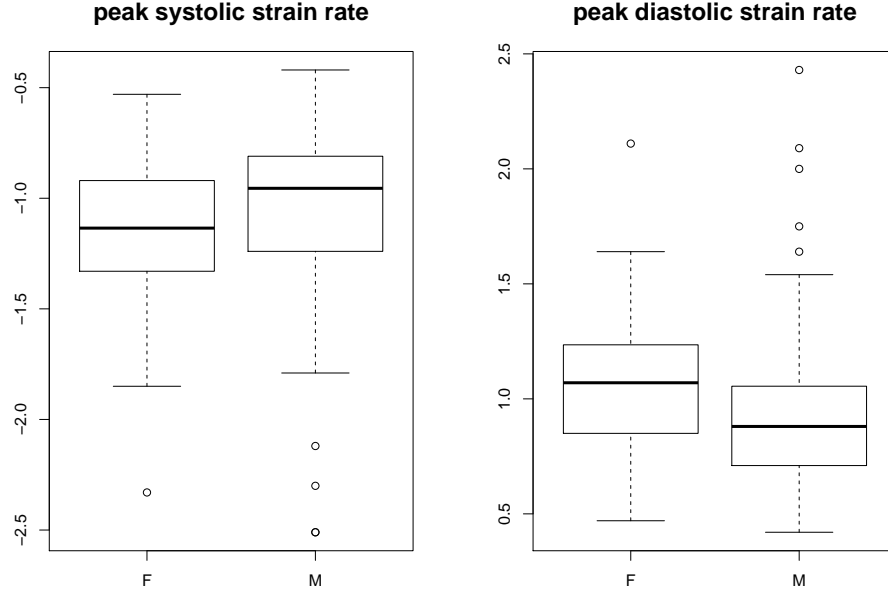


Figure 1: Boxplots of the systolic and diastolic peak strain rate for female and male patients.

6 Application: Analysis of the Data Example

As a data example, we consider cardiological measurements in the left ventricle of 188 healthy patients, that were recorded at the University clinic Ulm, Germany. Variables of interest are the peak systolic and diastolic strain rate (PSSR and PDSR, respectively), measured in circumferential direction, the end systolic and diastolic volume (ESV and EDV, respectively) as well as the stroke volume (SV). Note that the empirical covariance matrix is singular in this example, since stroke volume is calculated as the difference between end diastolic volume and end systolic volume. The empirical covariance matrices can be found in Section ?? in the supplementary material. We consider a one-way layout analyzing the factor 'gender' (female vs. male). Some descriptive statistics of the measurements for this factor are displayed in Table 6. Boxplots of the systolic and diastolic measurements are in Figures 1 and 2, respectively.

Table 6: Descriptive statistics of the cardiology data. Volume measurements (EDV, ESV and SV) are in ml , peak strain rate measurements are in $1/sec$.

Gender	n	Mean					Sd				
		EDV	ESV	SV	PSSR	PDSR	EDV	ESV	SV	PSSR	PDSR
female	92	124.37	41.39	82.98	-1.16	1.07	25.25	13.38	15.77	0.32	0.30
male	96	157.02	54.98	102.04	-1.07	0.94	31.02	15.98	18.70	0.39	0.35

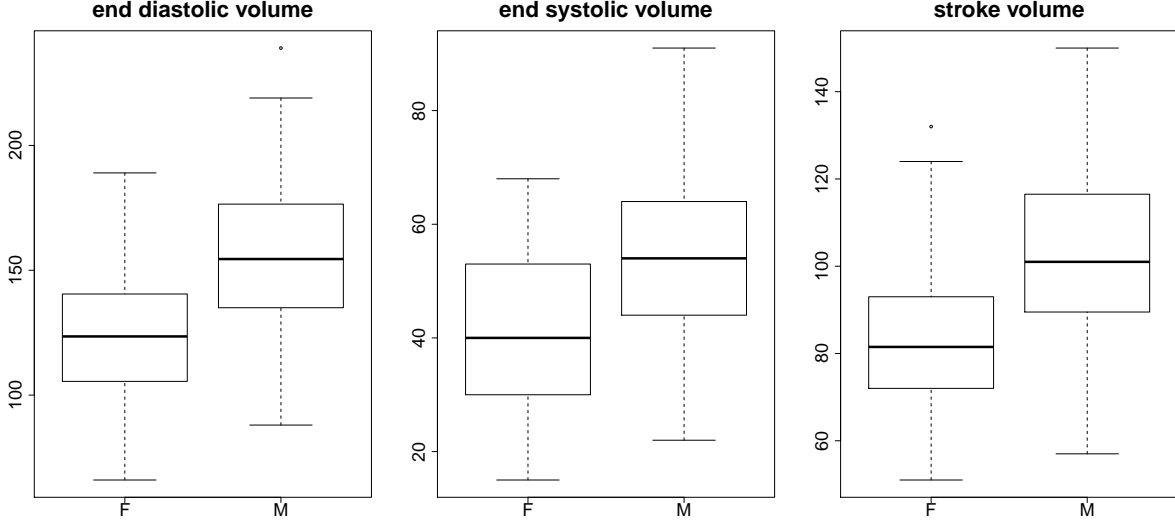


Figure 2: Boxplots of the end diastolic and systolic volume as well as stroke volume for female and male patients.

We now want to analyze, whether there is a significant difference in the multivariate means for female and male patients. The null hypothesis of interest thus is $H_0^{(1)} : \{(\mathbf{P}_2 \otimes \mathbf{I}_5)\boldsymbol{\mu} = \mathbf{0}\}$. Since the covariance matrix is singular in this example, we cannot apply the Wald-type test. Thus, we consider the parametric bootstrap approach of the MATS which yielded the best results in the simulation study. Computation of the MATS results in a value of $Q_N = 171.0011$ and the parametric bootstrap routine with 10,000 bootstrap runs gives a p -value of $p < 0.0001$, implying that there is indeed a significant effect of gender on the measurements.

In a second step we want to derive a confidence region for the factor 'Gender'. Here, we restrict our analyses to the strain rate measurements in order to be able to plot the confidence ellipsoids for the contrast of interest. That is, we consider the null hypothesis $H_0^{(2)} : \{\mathbf{T}\boldsymbol{\mu} = \mathbf{0}\} = \{\mu_{11} - \mu_{21} = \mu_{12} - \mu_{22}\} = \{\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}\}$, where μ_{ij} is the corresponding mean value of measurement j (systolic vs. diastolic measurement) in group i (female vs. male) and

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}.$$

The parametric bootstrap procedure with 10,000 bootstrap runs leads to a p -value of $p = 0.0146$ for the MATS, i.e., there is a significant effect of gender on the peak strain rate. We can now construct a confidence ellipsoid as described in Section 4.2 based on the parametric bootstrap quantile $c_{1-\alpha}^*$. Therefore, we need to compute the eigenvalues λ_j and eigenvectors $\mathbf{e}_j, j = 1, 2$ of $\mathbf{T}\hat{\mathbf{D}}_N\mathbf{T}$. The ellipse is centered at $\mathbf{T}\bar{\mathbf{X}}_{\bullet} = (-0.097, 0.126)'$. For the eigendecomposition of $\mathbf{T}\hat{\mathbf{D}}_N\mathbf{T}$, we obtain $\boldsymbol{\lambda} = (0.508, 0.412)$ as well as $\mathbf{e}_1 = (-1, 0)'$ and $\mathbf{e}_2 = (0, -1)'$, that is, the confidence ellipse extends $\sqrt{\lambda_1 \cdot c_{1-\alpha}^*/N} = 0.013$ units in the direction of \mathbf{e}_1 and 0.012 units in the direction of \mathbf{e}_2 . The corresponding ellipse is displayed in Figure 3. It turns out that female

patients have a lower systolic peak strain rate but a higher diastolic peak strain rate than male patients, a finding that is confirmed by the descriptive analysis in Table 6 and Figure 1.

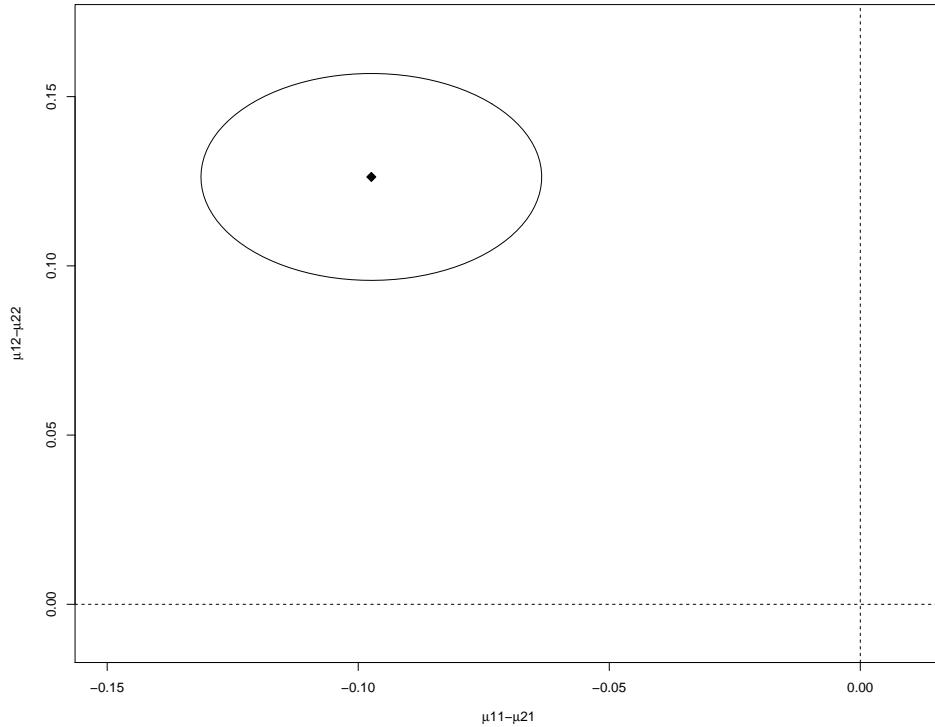


Figure 3: 95 % confidence ellipse for $\mu_1 - \mu_2$. The \blacklozenge denotes the center $T\overline{X}_\bullet$ of the ellipse. Female patients seem to have a lower systolic peak strain rate but a higher diastolic peak strain rate than male patients.

7 Conclusions and Discussion

We have investigated a test statistic for multivariate data (MATS) which is based on a modified Dempster statistic. Contrary to classical MANOVA models, we incorporate general heteroscedastic designs and allow for singular covariance matrices while postulating their existence as solely distributional assumption. Moreover, our proposed MATS statistic is invariant under linear transformations of the response variables.

In order to improve the small sample behavior of the test statistic, we have investigated two different bootstrap approaches, namely a parametric bootstrap and a wild bootstrap procedure. We have rigorously proven that both lead to asymptotically exact and consistent tests and even analyzed their local power behavior.

In a large simulation study, the parametric bootstrap turned out to perform very well in most scenarios, even with skewed data and heteroscedastic variances. Although the type-I error control is still not ideal in the latter case, the method performed advantageous over the parametric bootstrap of the WTS proposed in Konietschke et al. (2015) and has the additional advantage of being applicable to situations with singular covariance matrices. In situations with skewed distributions, the parametric bootstrap of the MATS yielded more robust results than the WTS. The wild bootstrap approach, in contrast, turned out to be very conservative for growing dimensions, especially in situations with negative pairing. We therefore recommend the parametric bootstrap based on the MATS.

Furthermore, we have constructed confidence regions and simultaneous confidence intervals for contrasts $c'\mu$ based on the bootstrap quantiles. These confidence regions provide an additional benefit for the analysis of multivariate data since they allow for more detailed insight into the nature of the estimates.

In order to facilitate application of the proposed methods, both the bootstrap tests and the calculation of confidence regions are implemented in the R-package **MANOVA.RM**.

Following the idea of Srivastava and Kubokawa (2013) we plan to extend our concepts to the high-dimensional setting, i.e., where the sample size N may be less than the dimension d . This approach looks promising, since we have seen in the simulation study that the MATS with the parametric bootstrap approach exhibited an improved type-I error control with increasing d . However, the extension to high-dimensional data requires different techniques and will be part of future research.

Acknowledgment

The authors would like to thank Dr. Jan Paul and Prof. Dr. Volker Rasche for providing the data example. This work was supported by the German Research Foundation projects DFG PA 2409/3-1 and PA 2409/4-1.

Supplementary Material to MATS: Inference for potentially Singular and Heteroscedastic MANOVA

Sarah Friedrich* and Markus Pauly*

Abstract

In this supplementary material to the authors' paper "MATS: Inference for potentially singular and heteroscedastic MANOVA" we provide the proofs of all theorems as well as some additional simulation results. Furthermore, we state the empirical covariance matrices of the data example and a simulation study mimicking this data set.

Keywords: MANOVA, Multivariate Data, Parametric Bootstrap, Confidence Regions, Singular Covariance Matrices

8 Proofs

Proof of Theorem 2.1

The result follows directly from the representation theorem for quadratic forms (Rao and Mitra, 1971) and the continuous mapping theorem by noting that $\sqrt{N}(\bar{\mathbf{X}}_{\bullet} - \boldsymbol{\mu})$ has, asymptotically, as $N \rightarrow \infty$, a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} = \lim_{N \rightarrow \infty} \boldsymbol{\Sigma}_N = \text{diag}(\kappa_i^{-1} \mathbf{V}_i)$. Moreover, $\hat{\mathbf{D}}_N$ is consistent for $\mathbf{D} = \text{diag}(\kappa_i^{-1} \sigma_{is}^2)$, where the latter is of full rank by assumption. Thus, $\mathbf{T} \hat{\mathbf{D}}_N \mathbf{T}$ converges in probability to $\mathbf{T} \mathbf{D} \mathbf{T}$ and since there is finally no rank jump in this convergence we eventually obtain

$$(\mathbf{T} \hat{\mathbf{D}}_N \mathbf{T})^+ \xrightarrow{P} (\mathbf{T} \mathbf{D} \mathbf{T})^+, \quad (8.1)$$

where \xrightarrow{P} denotes convergence in probability. \square

Proof of Theorem 3.1

Let

$$\mathbf{Y}_{ik} := \mathbf{X}_{ik} \cdot \mathbf{1}\{\|\mathbf{X}_{ik}\| \leq \delta\}$$

for $\delta > 0$. Then, \mathbf{Y}_{ik} has finite moments of any order, especially fourth moments exist. Analogously, given $\mathbf{Y} = (\mathbf{Y}_{ik})_{i,k}$, let $\mathbf{Y}_{ik}^* \stackrel{\text{i.i.d.}}{\sim} N(0, \hat{\mathbf{V}}_i(\delta))$, where $\hat{\mathbf{V}}_i(\delta) = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\mathbf{Y}_{ik} - \bar{\mathbf{Y}}_{i\cdot})(\mathbf{Y}_{ik} - \bar{\mathbf{Y}}_{i\cdot})'$ and define $\mathbf{Q}_{N,\delta}^* = N(\bar{\mathbf{Y}}_{\bullet}^*)' \mathbf{T} (\mathbf{T} \hat{\mathbf{D}}_{\delta}^* \mathbf{T})^+ \mathbf{T} \bar{\mathbf{Y}}_{\bullet}^*$, where $\hat{\mathbf{D}}_{\delta}^* = \text{diag}(\frac{N}{n_i} (\hat{\sigma}_{is}^*(\delta))^2)$ and $(\hat{\sigma}_{is}^*(\delta))^2$ is the empirical variance of Y_{iks}^* .

First, we apply the multivariate Lindeberg-Feller CLT to show that, given \mathbf{X} , $\sqrt{N} \bar{\mathbf{Y}}_{\bullet}^*$ converges in distribution to a normal distributed random variable. Therefore, consider $\widetilde{\mathbf{Y}}_i^* := \sqrt{N} \frac{1}{n_i} \mathbf{Y}_{ik}^*$ and $\mathbf{V}_i(\delta) = \text{Cov}(\mathbf{Y}_{ik})$. The Lindeberg-Feller CLT now yields convergence in distribution given the data \mathbf{X} to a normal distributed random variable

$$\sqrt{N} \bar{\mathbf{Y}}_{\bullet}^* \xrightarrow{\mathcal{D}|\mathbf{X}} N(\mathbf{0}, \boldsymbol{\Sigma}_{\delta}), \quad (8.2)$$

if we proof the following conditions:

$$\sum_{i=1}^a \sum_{k=1}^{n_i} E(\widetilde{\mathbf{Y}}_i^* | \mathbf{X}) = \mathbf{0} \quad (8.3)$$

$$\bigoplus_{i=1}^a \sum_{k=1}^{n_i} \text{Cov}(\widetilde{\mathbf{Y}}_i^* | \mathbf{X}) \xrightarrow{P} \boldsymbol{\Sigma}_{\delta} := \bigoplus_{i=1}^a \frac{1}{\kappa_i} \mathbf{V}_i(\delta) \quad (8.4)$$

$$\sum_{i=1}^a \sum_{k=1}^{n_i} E\left(\|\widetilde{\mathbf{Y}}_i^*\|^2 \cdot \mathbf{1}\{\|\widetilde{\mathbf{Y}}_i^*\| > \epsilon\} | \mathbf{X}\right) \xrightarrow{P} 0 \quad \forall \epsilon > 0. \quad (8.5)$$

Condition (8.3) follows since, given the data, $\mathbf{Y}_{ik}^* \stackrel{\text{i.i.d.}}{\sim} N(0, \widehat{\mathbf{V}}_i(\delta))$. For condition (8.4), note that

$$\bigoplus_{i=1}^a \sum_{k=1}^{n_i} \text{Cov}(\widetilde{\mathbf{Y}}_i^* | \mathbf{X}) = \bigoplus_{i=1}^a \sum_{k=1}^{n_i} \text{Cov} \left(\frac{\sqrt{N}}{n_i} \mathbf{Y}_{ik}^* | \mathbf{X} \right) = \bigoplus_{i=1}^a \sum_{k=1}^{n_i} \frac{N}{n_i^2} \widehat{\mathbf{V}}_i(\delta) = \bigoplus_{i=1}^a \frac{N}{n_i} \widehat{\mathbf{V}}_i(\delta).$$

Thus, since $\widehat{\mathbf{V}}_i(\delta)$ is a consistent estimator of $\mathbf{V}_i(\delta)$, (8.4) follows. For (8.5) note that

$$\mathbb{1}\{\|\widetilde{\mathbf{Y}}_i^*\| > \epsilon\} = 1 \Leftrightarrow \delta \geq \|\mathbf{Y}_{ik}^*\| > \frac{n_i}{\sqrt{N}}\epsilon = \frac{n_i}{N}\sqrt{N}\epsilon.$$

Since $\frac{n_i}{N} \rightarrow \kappa_i > 0$, the right hand side converges to infinity as $N \rightarrow \infty$. Therefore, for arbitrary fixed $\delta > 0$ and $\epsilon > 0$ we finally have $\mathbb{1}\{\|\widetilde{\mathbf{Y}}_i^*\| > \epsilon\} = 0$ for N large enough and Equation (8.5) follows. Altogether this proves (8.2).

Since $\widehat{\mathbf{D}}_\delta^* \xrightarrow{P} \mathbf{D}_\delta = \text{diag}(\frac{1}{\kappa_i} \text{Var}(Y_{iks})), i = 1, \dots, a, s = 1, \dots, d$ due to existence of finite fourth moments of the truncated random variables \mathbf{Y} , it now follows from continuous mapping and the representation theorem for quadratic forms that

$$Q_{N,\delta}^* = N(\overline{\mathbf{Y}}_\bullet^*)' \mathbf{T}(\mathbf{T}\widehat{\mathbf{D}}_\delta^* \mathbf{T})^+ \mathbf{T}\overline{\mathbf{Y}}_\bullet^* \xrightarrow{\mathcal{D}|\mathbf{X}} \tilde{Z}, N \rightarrow \infty,$$

in probability, see, e.g., Konietzschke et al. (2015) and the references cited therein for similar arguments. Here, $\tilde{Z} = \sum_{i=1}^a \sum_{s=1}^d \tilde{\lambda}_{is} \tilde{Z}_{is}$ with $\tilde{Z}_{is} \sim \chi_1^2$ and $\tilde{\lambda}_{is}$ are the eigenvalues of $\mathbf{T}(\mathbf{T}\mathbf{D}_\delta \mathbf{T})^+ \mathbf{T}\Sigma_\delta$. Furthermore, since

$$\text{Cov}(\mathbf{Y}_{ik}) = \text{Cov}(\mathbf{X}_{ik} \cdot \mathbb{1}\{\|\mathbf{X}_{ik}\| \leq \delta\}) \rightarrow \text{Cov}(\mathbf{X}_{ik}), \delta \rightarrow \infty,$$

by dominated convergence and analogously $\mathbf{D}_\delta \rightarrow \mathbf{D}, \delta \rightarrow \infty$, it follows that

$$\tilde{Z} \xrightarrow{\mathcal{D}|\mathbf{X}} Z, \delta \rightarrow \infty$$

in probability, where Z is the limit variable of Q_N given in Theorem 2.1. Thus, it remains to show that

$$\lim_{\delta \rightarrow \infty} \limsup_{N \rightarrow \infty} P(|Q_{N,\delta}^* - Q_N^*| > \epsilon | \mathbf{X}) \stackrel{\text{a.s.}}{=} 0 \text{ for all } \epsilon > 0.$$

Let $\widetilde{\mathbf{Y}} := \sqrt{N} \mathbf{T} \overline{\mathbf{Y}}_\bullet^*$, $\widetilde{\mathbf{X}} := \sqrt{N} \mathbf{T} \overline{\mathbf{X}}_\bullet^*$, $\mathbf{M}_\delta := (\mathbf{T} \mathbf{D}_\delta^* \mathbf{T})^+$ and $\mathbf{M} := (\mathbf{T} \mathbf{D}^* \mathbf{T})^+$. Then, $Q_{N,\delta}^* = \widetilde{\mathbf{Y}}' \mathbf{M}_\delta \widetilde{\mathbf{Y}}$ and $Q_N^* = \widetilde{\mathbf{X}}' \mathbf{M} \widetilde{\mathbf{X}}$ and therefore

$$Q_{N,\delta}^* - Q_N^* = \underbrace{\widetilde{\mathbf{Y}}' \mathbf{M}_\delta (\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}})}_{(A)} + \underbrace{(\widetilde{\mathbf{Y}}' \mathbf{M}_\delta - \widetilde{\mathbf{X}}' \mathbf{M}) \widetilde{\mathbf{X}}}_{(B)}.$$

First, consider part (A) and let $\boldsymbol{\xi}_{ik} := \mathbf{X}_{ik} - \mathbf{Y}_{ik} = \mathbf{X}_{ik} \mathbb{1}\{\|\mathbf{X}_{ik}\| > \delta\}$. Another application of the multivariate Lindeberg-Feller CLT shows that

$$\sqrt{N}(\overline{\mathbf{Y}}_\bullet^* - \overline{\mathbf{X}}_\bullet^*) \xrightarrow{\mathcal{D}|\mathbf{X}} N(\mathbf{0}, \widetilde{\Sigma}_\delta)$$

in probability, where $\widetilde{\Sigma}_\delta := \bigoplus_{i=1}^a \frac{1}{\kappa_i} \text{Cov}(\mathbf{X}_{ik} \mathbb{1}\{\|\mathbf{X}_{ik}\| > \delta\}) = \bigoplus_{i=1}^a \frac{1}{\kappa_i} \text{Cov}(\boldsymbol{\xi}_{ik})$. Thus,

$$\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}} \xrightarrow{\mathcal{D}|\mathbf{X}} N(0, \mathbf{T}\widetilde{\Sigma}_\delta)$$

in probability and the representation theorem again yields $\widetilde{\mathbf{Y}}' \mathbf{M}_\delta (\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}) \xrightarrow{\mathcal{D}|\mathbf{X}} B_\delta = \sum_{i=1}^a \sum_{s=1}^d \eta_{is}^{(\delta)} B_{is}^2$ in probability, where $B_{is}^2 \sim \chi_1^2$ and $\eta_{is}^{(\delta)}$ are the eigenvalues of $(\mathbf{T}\Sigma_\delta)^{1/2} \mathbf{M}_\delta (\mathbf{T}\widetilde{\Sigma}_\delta)^{1/2}$.

By dominated convergence it follows that $\widetilde{\Sigma}_\delta \rightarrow 0$ for $\delta \rightarrow \infty$. Since $\Sigma_\delta \rightarrow \Sigma$ and $\mathbf{D}_\delta \rightarrow \mathbf{D}$ we finally obtain $B_\delta \rightarrow 0$ as $\delta \rightarrow \infty$. Altogether, this proves

$$\lim_{\delta \rightarrow \infty} \limsup_{N \rightarrow \infty} (A) = 0.$$

It remains to consider part (B) which we expand as

$$(\widetilde{\mathbf{Y}}' \mathbf{M}_\delta - \widetilde{\mathbf{X}}' \mathbf{M}) \widetilde{\mathbf{X}} = [\widetilde{\mathbf{Y}}' \mathbf{M}_\delta - \widetilde{\mathbf{X}}' \mathbf{M}_\delta + \widetilde{\mathbf{X}}' (\mathbf{M}_\delta - \mathbf{M})] \widetilde{\mathbf{X}}.$$

Using similar arguments as above, it follows that given the data

$$(\widetilde{\mathbf{Y}}' - \widetilde{\mathbf{X}}') \mathbf{M}_\delta \widetilde{\mathbf{X}}$$

converges to 0 in probability for $N \rightarrow \infty$ and, subsequently, $\delta \rightarrow \infty$.

Finally, $(\hat{\sigma}_{is}^*(\delta))^2 - (\hat{\sigma}_{is}^*)^2$ converges to zero (where $(\hat{\sigma}_{is}^*)^2$ is the empirical variance of X_{iks}^*) by dominated convergence and consistency of the variance estimators and it follows that $\mathbf{M}_\delta - \mathbf{M}$ converges to 0. This concludes the proof. \square

Proof of Theorem 3.2

Analogous to the proof of Theorem 3.1, we define

$$\mathbf{Y}_{ik} := \mathbf{X}_{ik} \cdot \mathbb{1}\{\|\mathbf{X}_{ik}\| \leq \delta\}$$

for $\delta > 0$ as well as, given \mathbf{Y} , $\mathbf{Y}_{ik}^* = W_{ik}(\mathbf{Y}_{ik} - \overline{\mathbf{Y}}_{i\cdot})$ and $Q_{N,\delta}^* = N(\overline{\mathbf{Y}}^*)' \mathbf{T}(\mathbf{T}\widehat{\mathbf{D}}_\delta^* \mathbf{T})^+ \mathbf{T}\overline{\mathbf{Y}}^*$.

The first part of the proof follows analogous to the proof of Theorem 3.1 above. It remains to show that $(\hat{\sigma}_{is}^*(\delta))^2 - (\hat{\sigma}_{is}^*)^2$ converges to zero. Therefore, consider

$$(*) := (\hat{\sigma}_{is}^*(\delta))^2 - (\hat{\sigma}_{is}^*)^2 = \frac{1}{n_i} \sum_{k=1}^{n_i} W_{ik}^2 \xi_{iks}^2 - (\overline{Y}_{i\cdot}^*)^2 + (\overline{X}_{i\cdot}^*)^2,$$

where again $\xi_{iks} := X_{iks} - Y_{iks}$. For the first summand on the right hand side it holds

$$E\left(\frac{1}{n_i} \sum_{k=1}^{n_i} W_{ik}^2 \xi_{iks}^2 | \mathbf{X}\right) = \frac{1}{n_i} \sum_{k=1}^{n_i} \xi_{iks}^2 \xrightarrow[N \rightarrow \infty]{\text{a.s.}} E(X_{i1s}^2 \mathbb{1}\{\|\mathbf{X}_{ik}\| > \delta\}).$$

Now letting $\delta \rightarrow \infty$, it follows from dominated convergence that

$$E(X_{i1s}^2 \mathbb{1}\{\|\mathbf{X}_{ik}\| > \delta\}) \xrightarrow{\delta \rightarrow \infty} 0.$$

Concerning $(\bar{X}_{i.s}^*)^2 - (\bar{Y}_{i.s}^*)^2$, we first consider $\bar{X}_{i.s}^* - \bar{Y}_{i.s}^* = \frac{1}{n_i} \sum_{k=1}^{n_i} W_{ik} \xi_{iks}$. It holds that $E(\bar{X}_{i.s}^* - \bar{Y}_{i.s}^* | \mathbf{X}) = 0$ as well as $\text{Var}(\bar{X}_{i.s}^* - \bar{Y}_{i.s}^* | \mathbf{X}) \xrightarrow{N \rightarrow \infty} \text{Var}(\bar{X}_{i.s} - \bar{Y}_{i.s}) \xrightarrow{\delta \rightarrow \infty} 0$ and therefore

$$\lim_{\delta \rightarrow \infty} \limsup_{N \rightarrow \infty} P(|\bar{X}_{i.s}^* - \bar{Y}_{i.s}^*| > \epsilon | \mathbf{X}) = 0.$$

The continuous mapping theorem now implies

$$\lim_{\delta \rightarrow \infty} \limsup_{N \rightarrow \infty} P((\bar{X}_{i.s}^* - \bar{Y}_{i.s}^*)^2 > \epsilon | \mathbf{X}) = 0$$

and furthermore

$$(\bar{X}_{i.s}^* - \bar{Y}_{i.s}^*)^2 \mathbb{1}\{\bar{X}_{i.s}^* > \bar{Y}_{i.s}^*\} \xrightarrow{P} 0.$$

Therefore,

$$0 \leq ((\bar{X}_{i.s}^*)^2 - (\bar{Y}_{i.s}^*)^2) \mathbb{1}\{\bar{X}_{i.s}^* > \bar{Y}_{i.s}^*\} \rightarrow 0$$

and analogously

$$0 \leq ((\bar{X}_{i.s}^*)^2 - (\bar{Y}_{i.s}^*)^2) \mathbb{1}\{\bar{X}_{i.s}^* < \bar{Y}_{i.s}^*\} \rightarrow 0$$

and therefore

$$\lim_{\delta \rightarrow \infty} \limsup_{N \rightarrow \infty} P(|(\bar{X}_{i.s}^*)^2 - (\bar{Y}_{i.s}^*)^2| > \epsilon | \mathbf{X}) = 0.$$

Altogether, this implies by the general Markov inequality that

$$\lim_{\delta \rightarrow \infty} \limsup_{N \rightarrow \infty} P((*) > \epsilon | \mathbf{X}) \leq \lim_{\delta \rightarrow \infty} \limsup_{N \rightarrow \infty} \frac{1}{\epsilon} E((*) | \mathbf{X}) \stackrel{\text{a.s.}}{=} 0,$$

which concludes the proof. \square

Proofs of the results in Section 4.1

Theorems 3.1 and 3.2 directly imply that the corresponding bootstrap tests $\varphi^* = \mathbb{1}\{Q_N > c_{1-\alpha}^*\}$ asymptotically keep the pre-assigned level α , since $c_{1-\alpha}^*$ is the $(1 - \alpha)$ -quantile of the (conditional) bootstrap distribution, which, given the data, converges weakly to the null distribution of Q_N in probability.

For local alternatives $H_1 : \mathbf{T}\boldsymbol{\mu} = \frac{1}{\sqrt{N}}\mathbf{T}\boldsymbol{\nu}$, $\boldsymbol{\nu} \in \mathbb{R}^{ad}$, it holds that

$$\sqrt{N}\mathbf{T}\bar{\mathbf{X}}_{\bullet} = \sqrt{N}\mathbf{T}(\bar{\mathbf{X}}_{\bullet} - \boldsymbol{\mu}) + \mathbf{T}\boldsymbol{\nu}$$

has, asymptotically, as $N \rightarrow \infty$, a multivariate normal distribution with mean $\mathbf{T}\boldsymbol{\nu}$ and covariance matrix $\mathbf{T}\boldsymbol{\Sigma}\mathbf{T}$ and thus Q_N converges to $\zeta(\mathbf{T}\mathbf{D}\mathbf{T})^+\zeta$, where $\zeta \sim N(\mathbf{T}\boldsymbol{\nu}, \mathbf{T}\boldsymbol{\Sigma}\mathbf{T})$, by using (8.1) again. Thus, Theorems 3.1 and 3.2 imply that the bootstrap tests have the same asymptotic power as $\varphi_N = \mathbb{1}\{Q_N > c_{1-\alpha}\}$ and the asymptotic relative efficiency of the bootstrap tests φ^* compared to φ_N is 1 in this situation. \square

9 Further simulation results

9.1 One-way layout

Further simulation results for the one-way layout are displayed in Tables 7 – 9 for the lognormal, t_3 and double-exponential distribution, respectively. The parametric bootstrap of the MATS keeps the pre-assigned α -level very well for the t_3 and the double-exponential distribution. Note that the validity of the parametric bootstrap of the WTS has not yet been proven for the t_3 distribution, since fourth moments do not exist in this case. With lognormally distributed data and negative pairing (setting 3 and 4 with $\mathbf{n} = (20, 10)$), all procedures show a liberal behavior. For $d = 8$ dimensions, the wild bootstrap MATS behaves best in these scenarios, but the parametric bootstrap is better in almost all other scenarios. We therefore recommend the parametric bootstrap procedure.

9.2 Two-way layout

The simulation results for factor A and B are in Tables 10 and 11, respectively. The parametric bootstrap of the MATS performs best in most scenarios, whereas the wild bootstrap is slightly liberal in many settings.

10 Covariance matrices

For convenience, we included the estimated covariance matrices of the data example for female and male patients and the five outcome variables end diastolic volume (EDV), end systolic volume (ESV), stroke volume (SV), peak systolic strain rate (PSSR) and peak diastolic strain rate (PDSR) in this section. Covariance matrices, rounded to three decimals, for female patients (EDV, ESV, SV, PSSR, PDSR):

$$\begin{pmatrix} 6.931 & 3.087 & 3.844 & 0.022 & -0.012 \\ 3.087 & 1.947 & 1.140 & 0.018 & -0.013 \\ 3.844 & 1.140 & 2.704 & 0.004 & 0.002 \\ 0.022 & 0.018 & 0.004 & 0.001 & -0.001 \\ -0.012 & -0.013 & 0.002 & -0.001 & 0.001 \end{pmatrix}$$

and male patients:

$$\begin{pmatrix} 10.025 & 4.522 & 5.503 & 0.043 & -0.021 \\ 4.522 & 2.661 & 1.862 & 0.027 & -0.018 \\ 5.503 & 1.862 & 3.641 & 0.016 & -0.003 \\ 0.043 & 0.027 & 0.016 & 0.002 & -0.001 \\ -0.021 & -0.018 & -0.003 & -0.001 & 0.001 \end{pmatrix}.$$

10.1 Simulation study: Data example

Furthermore, we conducted a simulation study in order to analyze the behavior of our methods on data similar to the cardiology data. We therefore simulated $a = 2$ groups that correspond to the factor gender with $\mathbf{n} = (92, 96)$ and $d = 5$ dimensions and used the estimated covariance matrices from the data for simulation. We considered different distributions for the random errors and applied the parametric bootstrap to both WTS and MATS. Because of the rather unreliable results of the wild bootstrap, this method was not considered here. The results for $\text{nsim} = 5000$ simulation runs with $\text{nboot} = 5000$ bootstraps each are displayed in Table 12. The type-I error rates are similar for both approaches, but since the covariance matrices in this example are singular, the WTS should not be applied here.

Table 12: Type-I errors in % for the simulation study that mimics the data example: Compared are the parametric bootstrap versions of the WTS and the MATS.

	WTS (PBS)	MATS (PBS)
normal	4.4	4.5
χ_3^2	5.0	5.1
dexp	4.9	5.0
t_3	5.2	5.6
lognormal	4.4	4.2

References

- Bartlett, M. S. (1939). A note on tests of significance in multivariate analysis. *Mathematical Proceedings of the Cambridge Philosophical Society*, 35(02):180–185. Cambridge University Press.
- Bathke, A. C., Harrar, S. W., and Madden, L. V. (2008). How to compare small multivariate samples using nonparametric tests. *Computational Statistics & Data Analysis*, 52(11):4951–4965.
- Beyersmann, J., Termini, S. D., and Pauly, M. (2013). Weak convergence of the wild bootstrap for the Aalen–Johansen estimator of the cumulative incidence function of a competing risk. *Scandinavian Journal of Statistics*, 40(3):387–402.
- Brunner, E. (2001). Asymptotic and approximate analysis of repeated measures designs under heteroscedasticity. *Mathematical Statistics with Applications in Biometry*.
- Brunner, E., Konietschke, F., Pauly, M., and Puri, M. L. (2016). Rank-based procedures in factorial designs: hypotheses about non-parametric treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

- Brunner, E., Munzel, U., and Puri, M. L. (1999). Rank-score tests in factorial designs with repeated measures. *Journal of Multivariate Analysis*, 70(2):286–317.
- Brunner, E. and Puri, M. L. (2001). Nonparametric methods in factorial designs. *Statistical papers*, 42(1):1–52.
- Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427.
- Cameron, A. C. and Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.
- Chung, E. and Romano, J. P. (2016). Multivariate and multiple permutation tests. *Journal of Econometrics*, 193(1):76–91.
- Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, 146(1):162–169.
- Dempster, A. P. (1958). A high dimensional two sample significance test. *The Annals of Mathematical Statistics*, 29(4):995–1010.
- Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, 16(1):41–50.
- Friedrich, S., Brunner, E., and Pauly, M. (2017). Permuting longitudinal data in spite of the dependencies. *Journal of Multivariate Analysis*, 153:255–265.
- Friedrich, S., Konietschke, F., and Pauly, M. (2016). A wild bootstrap approach for nonparametric repeated measurements. *Computational Statistics & Data Analysis*.
- Harrar, S. W. and Bathke, A. C. (2012). A modified two-factor multivariate analysis of variance: asymptotics and small sample approximations. *Annals of the Institute of Statistical Mathematics*, 64(1):135–165.
- Hasler, M. and Hothorn, L. A. (2008). Multiple contrast tests in the presence of heteroscedasticity. *Biometrical Journal*, 50(5):793–800.
- Hotelling, H. (1951). A generalized t -test and measure of multivariate dispersion. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. The Regents of the University of California.
- Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3):346–363.
- Johnson, R. A. and Wichern, D. W. (2007). *Applied multivariate statistical analysis*. 6th edition, Prentice Hall.

- Konietschke, F., Bathke, A., Harrar, S., and Pauly, M. (2015). Parametric and nonparametric bootstrap methods for general MANOVA. *Journal of Multivariate Analysis*, 140:291–301.
- Konietschke, F., Hothorn, L. A., and Brunner, E. (2012). Rank-based multiple test procedures and simultaneous confidence intervals. *Electronic Journal of Statistics*, 6:738–759.
- Krishnamoorthy, K. and Lu, F. (2010). A parametric bootstrap solution to the MANOVA under heteroscedasticity. *Journal of Statistical Computation and Simulation*, 80(8):873–887.
- Lawley, D. (1938). A generalization of fisher’s z test. *Biometrika*, 30(1-2):180–187.
- Lin, D. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine*, 16(8):901–910.
- Liu, C., Bathke, A. C., and Harrar, S. W. (2011). A nonparametric version of wilks’ lambda - asymptotic results and small sample approximations. *Statistics & Probability Letters*, 81(10):1502–1506.
- Liu, R. Y. (1988). Bootstrap procedures under some non-iid models. *The Annals of Statistics*, 16(4):1696–1708.
- Mammen, E. (1993). *When does bootstrap work? Asymptotic results and simulations*. Springer Science & Business Media.
- Marcus, R., Eric, P., and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3):655–660.
- Pauly, M., Ellenberger, D., and Brunner, E. (2015). Analysis of high-dimensional one group repeated measures designs. *Statistics*, 49:1243–1261.
- Pesarin, F. and Salmaso, L. (2010). *Permutation tests for complex data: theory, applications and software*. John Wiley & Sons.
- Pesarin, F. and Salmaso, L. (2012). A review and some new results on permutation testing for multivariate problems. *Statistics and Computing*, 22(2):639–646.
- Pillai, K. (1955). Some new test criteria in multivariate analysis. *The Annals of Mathematical Statistics*, 26(1):117–121.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rao, C. and Mitra, S. (1971). *Generalized inverse of matrices and its applications*. Wiley New York.
- Smaga, Ł. (2016). Bootstrap methods for multivariate hypothesis testing. *Communications in Statistics-Simulation and Computation*. Just accepted.

- Sonnemann, E. (2008). General solutions to multiple testing problems. *Biometrical Journal*, 50(5):641–656.
- Srivastava, M. S. and Kubokawa, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis*, 115:204–216.
- Vallejo, G. and Ato, M. (2012). Robust tests for multivariate factorial designs under heteroscedasticity. *Behavior research methods*, 44(2):471–489.
- Vallejo, G., Fernández, M., and Livacic-Rojas, P. E. (2010). Analysis of unbalanced factorial designs with heteroscedastic data. *Journal of Statistical Computation and Simulation*, 80(1):75–88.
- Van Aelst, S. and Willems, G. (2011). Robust and efficient one-way MANOVA tests. *Journal of the American Statistical Association*, 106(494):706–718.
- Van Aelst, S. and Willems, G. (2013). Fast and robust bootstrap for multivariate inference: the R package FRB. *Journal of Statistical Software*, 53(3):1–32.
- Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 17(3):257–281.
- Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14(4):1261–1295.
- Xu, L.-W., Yang, F.-Q., Qin, S., et al. (2013). A parametric bootstrap approach for two-way ANOVA in presence of possible interactions with unequal variances. *Journal of Multivariate Analysis*, 115:172–180.

Table 7: Type-I error rates in % (nominal level $\alpha = 5\%$) for the parametric bootstrap (PBS) of the WTS and the MATS, as well as the wild bootstrap of the MATS in the one-way layout for the log-normal distribution.

d	Cov	n	WTS (PBS)	MATS (wild)	MATS (PBS)
$d = 4$	S1	(10, 10)	2.0	4.8	3.5
		(10, 20)	4.4	7.8	5.3
		(20, 10)	4.3	7.6	5.5
		(20, 20)	3.5	7.0	4.0
	S2	(10, 10)	1.9	4.3	3.2
		(10, 20)	4.5	7.4	4.9
		(20, 10)	4.4	7.5	5.4
		(20, 20)	3.4	6.9	3.9
	S3	(10, 10)	6.8	6.9	6.1
		(10, 20)	4.2	5.3	3.6
		(20, 10)	14.9	14.8	13.8
		(20, 20)	8.4	9.2	6.7
	S4	(10, 10)	7.0	6.8	6.2
		(10, 20)	4.2	5.0	3.6
		(20, 10)	15.4	14.4	13.5
		(20, 20)	8.7	9.3	6.8
$d = 8$	S1	(10, 10)	2.9	2.1	4.1
		(10, 20)	3.9	6.5	5.6
		(20, 10)	4.2	6.3	5.3
		(20, 20)	3.0	6.5	4.0
	S2	(10, 10)	2.8	0.7	2.5
		(10, 20)	4.0	4.1	4.8
		(20, 10)	4.1	3.6	4.1
		(20, 20)	3.0	4.3	3.1
	S3	(10, 10)	6.3	2.7	6.0
		(10, 20)	3.5	3.2	3.6
		(20, 10)	15.0	7.8	12.7
		(20, 20)	7.7	7.3	6.7
	S4	(10, 10)	7.0	1.6	6.0
		(10, 20)	3.6	1.7	2.8
		(20, 10)	15.5	6.7	13.6
		(20, 20)	8.0	6.3	7.0

Table 8: Type-I error rates in % (nominal level $\alpha = 5\%$) for the parametric bootstrap (PBS) of the WTS and the MATS, as well as the wild bootstrap of the MATS in the one-way layout for the t_3 distribution.

d	Cov	n	WTS (PBS)	MATS (wild)	MATS (PBS)
$d = 4$	S1	(10, 10)	3.6	5.8	4.4
		(10, 20)	4.8	6.7	4.8
		(20, 10)	3.9	6.2	4.3
		(20, 20)	4.0	7.2	4.5
	S2	(10, 10)	3.7	5.4	4.1
		(10, 20)	4.7	6.4	4.6
		(20, 10)	4.1	6.0	4.2
		(20, 20)	4.0	7.2	4.3
	S3	(10, 10)	4.1	3.2	3.1
		(10, 20)	4.3	5.4	4.2
		(20, 10)	4.9	3.1	3.3
		(20, 20)	3.9	4.9	3.8
	S4	(10, 10)	4.2	3.2	3.1
		(10, 20)	4.4	4.9	3.9
		(20, 10)	4.9	3.0	3.2
		(20, 20)	4.0	4.9	4.0
$d = 8$	S1	(10, 10)	3.5	2.9	4.7
		(10, 20)	4.8	4.9	4.7
		(20, 10)	5.0	4.2	4.2
		(20, 20)	3.9	7.1	4.7
	S2	(10, 10)	3.4	1.4	3.8
		(10, 20)	4.8	3.0	4.1
		(20, 10)	5.0	2.7	3.5
		(20, 20)	3.9	4.8	3.9
	S3	(10, 10)	5.2	0.8	3.2
		(10, 20)	3.5	3.0	4.2
		(20, 10)	8.6	0.6	2.7
		(20, 20)	4.0	3.0	3.6
	S4	(10, 10)	5.0	0.4	2.6
		(10, 20)	3.6	2.3	3.6
		(20, 10)	8.3	0.3	2.5
		(20, 20)	3.8	2.2	3.2

Table 9: Type-I error rates in % (nominal level $\alpha = 5\%$) for the parametric bootstrap (PBS) of the WTS and the MATS, as well as the wild bootstrap of the MATS in the one-way layout for the double-exponential distribution.

d	Cov	n	WTS (PBS)	MATS (wild)	MATS (PBS)
$d = 4$	S1	(10, 10)	4.1	5.5	4.5
		(10, 20)	5.0	6.6	4.8
		(20, 10)	4.1	5.7	4.2
		(20, 20)	5.1	8.0	5.6
	S2	(10, 10)	4.1	5.6	4.5
		(10, 20)	4.9	6.6	4.8
		(20, 10)	4.0	5.6	4.2
		(20, 20)	5.1	7.7	5.4
	S3	(10, 10)	4.5	3.1	3.3
		(10, 20)	4.3	5.1	4.4
		(20, 10)	5.1	3.0	3.2
		(20, 20)	4.7	5.4	4.9
	S4	(10, 10)	4.6	3.2	3.7
		(10, 20)	4.6	5.1	4.3
		(20, 10)	5.0	3.0	3.3
		(20, 20)	4.7	5.5	4.7
$d = 8$	S1	(10, 10)	3.4	2.2	4.6
		(10, 20)	5.2	4.7	4.8
		(20, 10)	4.7	4.7	4.9
		(20, 20)	4.3	6.7	4.7
	S2	(10, 10)	3.6	0.9	3.4
		(10, 20)	5.3	2.8	4.3
		(20, 10)	4.6	3.1	4.3
		(20, 20)	4.3	4.8	4.3
	S3	(10, 10)	4.8	0.3	3.0
		(10, 20)	4.1	2.9	4.2
		(20, 10)	8.1	0.8	3.0
		(20, 20)	5.0	2.8	3.6
	S4	(10, 10)	4.7	0.2	2.5
		(10, 20)	4.2	1.9	3.7
		(20, 10)	7.9	0.4	2.5
		(20, 20)	4.9	2.2	3.5

Table 10: Type-I error rates in % (nominal level $\alpha = 5\%$) for the parametric bootstrap (PBS) of the WTS and the MATS, as well as the wild bootstrap of the MATS when testing for an effect of factor A in a two-way layout with $d = 4$ dimensional observations.

distr	Cov	n	WTS (PBS)	MATS (wild)	MATS (PBS)
normal	S8	$\mathbf{n}^{(1)}$	4.6	7.1	4.8
		$\mathbf{n}^{(2)}$	4.9	7.4	5.3
		$\mathbf{n}^{(3)}$	5.0	6.9	5.1
		$\mathbf{n}^{(4)}$	4.9	8.0	5.2
	S9	$\mathbf{n}^{(1)}$	4.6	7.0	4.7
		$\mathbf{n}^{(2)}$	4.8	7.4	5.2
		$\mathbf{n}^{(3)}$	4.9	7.3	5.0
		$\mathbf{n}^{(4)}$	4.9	7.8	5.1
	S10	$\mathbf{n}^{(1)}$	4.6	4.4	4.3
		$\mathbf{n}^{(2)}$	4.5	5.3	4.9
		$\mathbf{n}^{(3)}$	5.4	4.3	4.2
		$\mathbf{n}^{(4)}$	5.0	5.6	5.0
	S11	$\mathbf{n}^{(1)}$	4.5	4.4	4.3
		$\mathbf{n}^{(2)}$	4.8	4.8	4.6
		$\mathbf{n}^{(3)}$	5.4	4.2	4.1
		$\mathbf{n}^{(4)}$	5.0	4.9	4.6
χ_3^2	S8	$\mathbf{n}^{(1)}$	3.8	6.8	4.6
		$\mathbf{n}^{(2)}$	4.5	6.9	4.7
		$\mathbf{n}^{(3)}$	5.0	6.9	4.9
		$\mathbf{n}^{(4)}$	4.3	7.6	4.5
	S9	$\mathbf{n}^{(1)}$	3.8	6.7	4.6
		$\mathbf{n}^{(2)}$	4.4	6.8	4.6
		$\mathbf{n}^{(3)}$	5.0	6.8	5.0
		$\mathbf{n}^{(4)}$	4.4	7.4	4.2
	S10	$\mathbf{n}^{(1)}$	5.0	5.5	4.8
		$\mathbf{n}^{(2)}$	4.2	4.8	4.0
		$\mathbf{n}^{(3)}$	6.8	6.2	5.8
		$\mathbf{n}^{(4)}$	5.0	5.5	4.6
	S11	$\mathbf{n}^{(1)}$	4.4	4.7	4.3
		$\mathbf{n}^{(2)}$	4.1	4.1	3.6
		$\mathbf{n}^{(3)}$	6.3	5.7	5.3
		$\mathbf{n}^{(4)}$	4.6	4.7	4.2

Table 11: Type-I error rates in % (nominal level $\alpha = 5\%$) for the parametric bootstrap (PBS) of the WTS and the MATS, as well as the wild bootstrap of the MATS when testing for an effect of factor B in a two-way layout with $d = 4$ dimensional observations.

distr	Cov	n	WTS (PBS)	MATS (wild)	MATS (PBS)
normal	S8	$\mathbf{n}^{(1)}$	4.5	6.6	4.9
		$\mathbf{n}^{(2)}$	5.2	7.2	5.3
		$\mathbf{n}^{(3)}$	5.3	6.5	4.7
		$\mathbf{n}^{(4)}$	4.8	7.8	5.2
	S9	$\mathbf{n}^{(1)}$	4.4	6.5	4.6
		$\mathbf{n}^{(2)}$	5.3	7.4	5.3
		$\mathbf{n}^{(3)}$	5.3	6.5	4.4
		$\mathbf{n}^{(4)}$	4.7	7.6	5.3
	S10	$\mathbf{n}^{(1)}$	4.8	4.6	4.3
		$\mathbf{n}^{(2)}$	5.2	5.6	5.1
		$\mathbf{n}^{(3)}$	5.9	4.8	4.8
		$\mathbf{n}^{(4)}$	4.5	5.2	4.9
	S11	$\mathbf{n}^{(1)}$	4.9	4.3	4.1
		$\mathbf{n}^{(2)}$	5.3	4.8	4.7
		$\mathbf{n}^{(3)}$	5.7	4.9	4.8
		$\mathbf{n}^{(4)}$	4.6	5.0	4.8
χ_3^2	S8	$\mathbf{n}^{(1)}$	4.1	6.8	4.6
		$\mathbf{n}^{(2)}$	4.4	6.4	4.4
		$\mathbf{n}^{(3)}$	4.5	6.4	4.3
		$\mathbf{n}^{(4)}$	5.0	8.2	5.2
	S9	$\mathbf{n}^{(1)}$	4.0	6.5	4.4
		$\mathbf{n}^{(2)}$	4.4	6.4	4.0
		$\mathbf{n}^{(3)}$	4.6	6.3	4.2
		$\mathbf{n}^{(4)}$	4.9	8.1	5.2
	S10	$\mathbf{n}^{(1)}$	4.4	4.4	4.0
		$\mathbf{n}^{(2)}$	4.1	4.9	3.9
		$\mathbf{n}^{(3)}$	5.2	4.5	4.2
		$\mathbf{n}^{(4)}$	5.2	5.8	4.9
	S11	$\mathbf{n}^{(1)}$	4.4	4.1	3.8
		$\mathbf{n}^{(2)}$	4.0	4.3	3.9
		$\mathbf{n}^{(3)}$	5.2	4.6	4.2
		$\mathbf{n}^{(4)}$	5.0	5.5	4.8